Assessing the latent linguistic structure of oral dream narratives

Assessing the latent linguistic structure of oral dream narratives

Abstract

The narrative structure of dream narratives is assumed to follow a fixed hierarchical order that reflects the temporal sequence of the original dream event. This corpus-based study aims to assess the latent linguistic structure and function of the salient linguistic features that that span the unit boundaries of orally elicited dream narratives (N = 39). A hierarchical cluster analysis identified that the content of orally elicited dream narratives is based on five linguistic clusters, including 1) spoken discourse style, 2) level of detail, 3) self- and other references, 4) reasoning processes, and 5) spatial and motion processes. The results are analysed by exploring the discursive function of the linguistic clusters within the dream narratives.

Keywords: Dream narratives, cluster analysis, content analysis, corpus-linguistics, latent linguistic structure.

1. Introduction

This study assesses the latent linguistic structure in orally elicited dream narratives employing a hierarchical cluster analysis. Dream narratives typically recapitulate events in the same temporal order as they occurred in the original, personal past event, resulting in a homogenous narrative structure (Labov & Waletzky, 1967). This narrative structure is assumed to follow a distinctive sequencing schema, resulting in a fixed temporal order of six narrative units: 1) Abstraction, 2) Orientation, 3) Complication, 4) Evaluation, 5) Resolution, and 6) Coda. Qualitative linguistic research has also shown that the contents of orally elicited dream narratives in British English follow a temporally linear structure similar to that of everyday personal narratives (Cariola, 2008). A cross-linguistic computerised content analysis assessing the latent linguistic structure of German and American English written dream diary entries identified three linguistic dimensions: 1) self- and other references, 2) level of description, and 3) spatial and motion processes; this analysis demonstrates that the latent linguistic structures did not vary substantially between the languages (Cariola, 2011). Surprisingly, computer-assisted linguistic research has not yet explored the latent linguistic structure of British English oral dream narratives.

Although it could be assumed that the linguistic differences between oral and written dream narratives would be relatively small, linguistic research has devoted much attention to investigating linguistic differences between the production of spoken and written text. Typically, written language is influenced by the structural constraints of producing a coherent text that conveys meaningful information. Spoken language, on the other hand, is limited by situational contexts that are susceptible to unplanned linguistic variables, such as repetitions, false starts, paralinguistic cues, and pauses (e.g., Brown & Yule, 1983; Halliday, 1994). Due to the contextual differences that govern oral and written language production, written narratives have been found to be more lexically and structurally complex but grammatically simpler than spoken texts, while spoken narratives are lexically and structurally simpler but grammatically more complex than written texts. Based on these identified differences that distinguish spoken from written text, it would be reasonable to expect that the grammatical complexity of verbally communicated dream narratives would yield a more intricate latent linguistic structure than written dream narratives, such as dream diary entries.

One of the earliest computer-assisted studies to assess linguistic dimensions and their discursive functions in different types of written and spoken discourse was conducted by Biber (1988). This was performed with the use of a computer-assisted analysis that gauged the frequencies of linguistic variables and a multi-dimensional (MD) analysis to explore "empirically and quantitatively linguistic features that co-occur in texts because they work together to mark some common underlying function within the text" (Biber, 1988, p.55). Based on this approach, five main factors were identified and ordered in decreasing variance. Subsequently, they were interpreted in relation to their functional dimensions, including the following: 1) *Information vs. Involved production*, 2) *Narrative vs. Non-narrative concerns*, 3) *Elaborated vs. Situation-dependent reference*, 4) *Overt expression of persuasion*, 5) *Abstract vs. Non-abstract information.*[1] The first dimension, 'Information vs. Involved production', represents the most basic language differences between spoken and written text. It differentiates between discourse reflecting an informational style that conveys specific meaning and carefully edited information (i.e., fiction, professional letters, and broadcasts) versus an involved style emphasising affective and immediate contextual processes that communicate generalised content (i.e., telephone conversations, personal letters, and spontaneous speeches). The second dimension, 'Narrative vs. Non-narrative concerns', identifies the way discourse uses exclusively present tense or past tense. It distinguishes between discourse with a focus on past events involving participants and the use of reported speech (i.e., fiction, biography, and spontaneous speeches) versus a non-narrative purpose, as found in expository and descriptive text (i.e., personal letters, interviews, and official documents). The third dimension, 'Elaborated vs. Situation-dependent reference', relates to discourse with specific identification of the referents in a text (i.e., telephone conversations, fiction, and personal letters) versus

[1] Biber (1988) outlined a total of seven linguistic dimensions; however, Biber and colleagues rejected the final two factors, and the subsequent literature proposes only five linguistic dimensions assumed to underlie spoken and written discourse (Biber, Conrad, & Reppen, 1998). The sixth dimension, 'Online informational elaboration', does not reflect a negative factor loading. Its function refers to discourse that is unplanned and text that is produced in highly constrained circumstances with loose and fragmented informational circumstances versus highly planned text with integrated and informative text. The seventh dimension, 'Academic hedging or qualifiers', is based on a weak factor loading and refers to the degree of probability of an assertion.

elaborated implicit deictics and references that are situated outside the text (i.e., spontaneous speeches, press reviews, and official documents). The fourth dimension, 'Overt expression of persuasion', relates to discourse conveying the narrator's own point of view (i.e., face-to-face conversations, press reviews, and broadcasts) versus an argumentative discourse style adopted to persuade the addressee (i.e., professional letters, public conversations, and personal letters). The fifth dimension, 'Abstract vs. Non-abstract information', differentiates between discourses with a formal style that emphasise abstract and technical information (i.e., press reviews, official documents, and professional letters) versus other types of discourse that communicate information that is not abstract (i.e., biographies, fiction, and telephone conversations).

Based on a similar multivariate approach, Pennebaker and King (1999) attempted to assess the linguistic structure of the written introspective accounts of college students. The analysis yielded four factors: 1) Immediacy, 2) Making Distinctions, 3) The Social Past, and 4) Rationalisation. The first factor, 'Immediacy', was based on positive loadings of first-person singular pronouns, discrepancy words and present tense words, and negative loadings of articles and words with more than 6 letters. The second factor, 'Making Distinctions', loaded positively for exclusion words, tentative words, and negation words, and loaded negatively for inclusion words. The third factor, 'The Social Past', comprised the positive loadings on past tense words and social words and the negative loadings on present tense words and positive emotions words. The fourth factor, 'Rationalisation', had positive loadings on insight words and causation words and a negative loading for negative emotion words.

This study aimed to identify the latent linguistic structure of orally elicited dream narratives. This was performed by employing a multivariate statistical method based on the assumption that "language is encoded and decoded as a combination of interrelated language features" (Mulac & Lundell, 1994, p. 300). Although Biber (1988) employed a multi-feature/multivariate approach to identify covarying linguistic features as a means of differentiating between registers or discourse types, the use of factor analysis might not represent an adequate method for the identification of linguistic dimensions within an a priori determined discourse type, such as, for example, dream narratives. Factor analysis has been shown to be appropriate for the exploration of distinctive linguistic dimensions and their

comparative functions as a means of differentiating discourse types. The use of a hierarchical cluster analysis, however, might represent a more suitable approach to exploring the latent linguistic structure that is based on the grouping of linguistic variables that assume similar functions within categorically exclusive discourse types. Factor analysis reduces a set of highly positively and negatively correlating linguistic variables into a set of manifest dimensions, whereas cluster analysis combines linguistic variables into groups based on their functional similarities, in which the linguistic variables in one cluster are assumed to be more similar to each other than to the variables in other cluster groupings. The use of cluster analysis in the assessment of similar linguistic functions has been employed, for example, in truthful and deceptive autobiographical memories (Cariola, O'Connor, & Conway, 2009) the cross-linguistic analysis of dream narratives (Cariola, 2011b).

2. Method

2.1 Participants

Forty-four undergraduate and postgraduate students, 21 male and 23 female, with a mean age of 25.63 years (ranging from 18 to 41 years) from London University were used for this study on oral dream narratives. Narratives of native British English speakers were considered for this study only, and therefore the narratives of five participants were excluded for further analysis. Participants took part voluntarily and were unknown to the researcher at the time of the study. They were approached in the foyer of the University library; for this reason, they represented opportunity samples.

2.2 Procedure

The participants were approached within library facilities. All participants signed a participant consent form and were informed that the purpose of the study was to investigate the content of dream narratives. All participants were given standardised instructions to complete a dream recall task; they were prompted to orally relate a recent and vivid dream memory in as much detail as possible with the question "Could you please recall a dream in as much detail as possible?" The researcher recorded all dream narratives with the use of a microphone and an iPod mini. The

participants were reassured that any personal information that could be linked to them would not be transcribed. After completing the experimental tasks, all participants were debriefed and thanked for participating.

2.3 Design

A hierarchical cluster analysis was conducted (Everitt, 1979) to assess the clustering of linguistic variables that form the latent linguistic structure of dream narratives. This was performed by standardising the data to a mean of 1 to reduce variances (Aldenderfer & Blashfield, 1984). Linguistic variables were clustered using the average linkage between-group algorithms (Soka & Sneath, 1963). The average linkage is based on the average distance between variables to compute their relative variable distances. The average linkage algorithm is considered highly accurate because it produces the greatest average similarity of clusters compared to the clusters of other algorithms by "maintaining the original similarity relations between entities" (Edelbrock, 1979, p.375). The average linkage, however, has also been criticised for forming non-conformist clusters of outliers when the sizes of the clusters increase (Blashfield, 1976; Williams, Clifford, & Lance, 1972).

2.4 Objective measures

All narratives were analysed with the Linguistic Inquiry and Word Count (LIWC 2007) text analysis program (Pennebaker, Francis, & Booth, 2001), which calculates the percentage of words and the lexical content of over 80 predefined linguistic categories, i.e., 4 descriptive variables (e.g., word count, words per sentence, and dictionary words), 22 grammatical variables (e.g., verbs, auxiliary verbs, and articles), 32 psychological construct variables, 7 personal concern variables (e.g., work, home, and leisure), 3 paralinguistic variables (e.g., fillers, assents, and non-fluencies), and 12 punctuation variables (e.g., commas, question marks, and colons).

2.6 Selection of linguistic variables

A preliminary analysis was conducted to select relevant linguistic variables for further analysis (Cariola, O'Connor, Conway, 2009; Cariola, 2011; Pennebaker & King,

1999). First, the linguistic variables with a mean frequency of ≤ 1% were excluded from further analysis. Second, linguistic variables were excluded if they co-correlated with other linguistic categories (see Pennbaker & King, 1999). Third, main dictionaries (e.g., negative emotions) were only included if their subdictionaries (e.g., anxiety, anger, and sadness) were below a mean frequency of 1%. Fourth, linguistic variables referring to the participants' personal concerns were excluded (e.g., achievements, leisure, and money). Fifth, descriptive variables, except word count, paralinguistic and punctuation variables, were excluded from further analysis. A total of seventeen linguistic variables were selected, i.e., first-person singular pronouns (e.g., I, me, and myself), third-person singular pronouns (e.g., she, her, and herself), past tense (e.g., went, ran, and had), present tense (e.g., is, does, and hear), adverbs (e.g., very, really, and quickly), quantifiers (e.g., few, many, and much), positive emotion words (e.g., love, nice, and sweet), negative emotions (e.g., hurt, ugly, and nasty), insight words (e.g., think, know, and consider), causation words (e.g., because, effect, and hence), tentative words (e.g., maybe, perhaps, and guess), certainty words (e.g., always and never), inclusion words (e.g., and, with, and include), exclusion words (e.g., but, without, and exclude), seeing words (e.g., view, saw, and seen), motion words (e.g., arrive, car, and go), and spatial words (e.g., down, in, and thin).

3. Results

3.1 Descriptive statistics

The descriptive statistics of the word counts and the seventeen linguistic variables are presented in Table 1. The text corpus has a total of 8,434 words with a mean of 216.26 per narrative (SD = 116.31). As for the linguistic variables, inclusion words (M = 9.17, SD = 3.83), first-person singular pronouns (M = 8.84, SD = 3.23), and past tense processes (M = 8.88, SD = 3.81) showed the highest frequencies, while seeing processes (M = 1.03, SD = 1.38), causation words (M = 1.06, SD = .90), and third-person singular pronouns (M = 1.86, SD = 1.26) had the lowest frequencies.

[Table 1 here]

3.2 Cluster analysis

A hierarchical cluster analysis produced a dendogram based on seventeen linguistic variables (Figure 1). The classification of clusters is typically determined by the arbitrary distinctness and compactness of a cluster's branch lengths along the scale along the top of the chart before it combines into a larger cluster (Holland, 2006). The greatest branch length for all clusters was established at the right cut-off point of 21 along the 0-25 x-axis; therefore, the dendogram proposes a 5-cluster solution. The first cluster contained 3 linguistic variables, i.e., adverbs, exclusion words and positive emotion words. The second cluster contained 4 linguistic variables, i.e., tentative words, certainty words, quantifiers, and present tense verbs. The third cluster contained 4 linguistic variables, i.e., first-person singular pronouns, seeing words, negative emotion words and third-person singular pronouns. The fourth cluster contained 3 linguistic variables, i.e., past tense verbs, insight words and causation words. The fifth cluster contained 3 linguistic variables, i.e., inclusion words, motion words, and spatial words.

[Figure 1 here]

3.3. Functional analysis of clusters

The dendogram represents a visual similarity matrix of the latent linguistic structure of dream narratives in which the cluster groupings of the linguistic variables typify the salient linguistic function. In order to interpret the discursive function of the clusters, it is important to assess the individual function of the linguistic variables that form the clusters.

Interpretation of the first cluster

The first cluster comprises the linguistic variables of adverbs, exclusion words and positive emotion words. Positive emotion words represent a category encompassing lexical items that convey positive affective concepts, achievements and social interactions in relation to the experience of positive events (Tausczik & Pennebaker, 2010; Kahn, Tobin, Massey, & Anderson, 2007). Adverbs are a word class that typically modifies the manner, time and place of events and actions. The generalised

adverbial markers (e.g., “just”, “well”, “really” and “basically”) assume the function of conversational fillers to maintain textual coherence (Schourup, 1999; Bieber, 1988). These markers express the standard adverbial meaning of “simply” and “only” and have been identified as linguistic innovations used most often by young people in contemporary urban speech communities (Tagliamonte, 2005). Additionally, adverbial markers are used as hedges and boosters in speech; adverbial hedges (e.g., “perhaps”, “maybe”, and “probably”) weaken a linguistic statement, expressing a sense of uncertainty and allowing the reader the option of disagreeing with the statement (Lakoff, 1975), whereas boosters (e.g., “very”, and “really”) purposefully intensify the statement to be believed and taken seriously (Holmes, 1990). The use of hedges and boosters is often perceived as a characteristic expression of a lack of confidence, but it might also indicate a politeness marker (Holmes, 1995). According to Wilson (2009), the use of hedges in spoken discourse also functions as an editing function of the real time composed dream narrative, resulting in a sequence of short independent clauses. Exclusion words negotiate the truth propositions of varying pieces of information, for example, through the use of adverbs (e.g., “rather”), conjunctions (e.g., “but”, “or”), and negations (e.g., “not“), as a means of indirectly approximating the accuracy of a statement (Izutsu, 2008; Fraser & Malamud-Makowski, 1996). In summary, the linguistic variables of the first cluster can be associated with a *spoken discourse style* that establishes textual coherence of the continuous processing of the dream information and minimises the overall impact of the spoken dream narrative. Taking the inherent sense of vagueness of recall of a dream event into consideration, this speech style might reflect a narrator’s politeness strategy in the truthful description of a dream narrative by mitigating the implicit boundaries of the coarse and detailed information being conveyed (Goldsmith et al., 2005; Goldsmith et al., 2002). It allows the narrator to communicate the broad details of a dream’s event actions and approximate an underlying realistic quality while maintaining vagueness about whether the event happened in the precise way that it is being recalled and recapitulated in the dream narrative, e.g., [4]:

“The dream <u>basically</u> involved one of my flatmates being pregnant which she isn't. And my other flatmate just…we were <u>just</u> hanging out <u>just</u> talking. That was <u>basically</u> it. And that was the only thing I can remember about the

dream was just her being pregnant. And we were just sitting around watching TV and we went out as well. But that was it".

Interpretation of the second cluster

The second cluster is based on the linguistic variables of present tense verbs, tentative words, quantifiers, and certainty words. The present tense typically represents and locates events in the present moment. Tentative words generally express a notion of uncertainty by negating the absolute truth proposition of a statement, for example, through the use of modifying adverbs (e.g., "maybe") and modal verbs (e.g., "might" or "may") that convey a low degree of commitment and thus downgrade the force and directness of the utterance (Holmes, 1995). Quantifiers (e.g., "none", "total", and "rest") express the general proportion of scope and the restriction of a specified entity and noun (McNally & Kennedy, 2008). The actual act of remembering the dream's content assumes an authentic act with an implicit certainty about the conveyed information (Ochs & Capps, 1997). Certainty words represent a distinctive class of modals (e.g., "must"), sentential modifiers (e.g., "necessarily" and "unfortunately"), temporal quantifiers, (e.g., "always", "ever", and "never") and universal quantifiers (e.g., "all"). Overall, this cluster appears to focus on the *level of detail* that negotiates the certainty and commitment of the descriptive details of the dream content series as they are retrieved from the narrator's knowledge base. Conversely, the narrator may also use the narrative present tense to generate a conversational immediacy and stress the notion of the narrator's vivid dream recall (Comrie, 1976), e.g., [3]:

"Okay so how did would it begin. I am just trying to remember it because I don't ever remember my dreams or any what. So maybe I was on top of a cliff well on a field on a hill. And then for some reason or another I am playing maybe football. And I end up hanging of the cliff though I am not scared or anything. But I am hanging off the cliff and like I could drop down to the ocean at any point although I don't. And then a complete well a guy I haven't seen for 15 years turns up on a bicycle says 'hello' and pulls me up and we go off and start playing football again. That is sort of the dream I can only remember. It seems quite short although in your mind you it felt for ages. But that is pretty much the dream I had".

The notion of uncertainty expressed through the use of tentative words, certainty words and quantifiers may also reflect the narrator's conscious awareness of his lack of retrieval accuracy. The discrepancy between the fuzzy vagueness of the conveyed dream content and the concurrent sense of awareness of its inaccuracy reconstruct the actual past dream event, resulting in a conflicting narrative style in which the narrators "establishe themselves as authorities and at the same time establish [themselves] as a liar or culprit" (Ochs & Capps, 1997). Although memories are inherently prone to suggestibility (Loftus, 1980), the notion of a fuzzy conscious awareness of an actual dream experience may relate to the Freudian principle of secondary revision (Freud, 1900). In such a secondary revision, the dynamic inhibiting forces of the ego selectively block and distort the recall of unacceptable manifest dream content upon awakening, restraining the preconscious from acknowledging these primitive and chaotic parts of one's unconscious knowledge, such as intense forces, motivations, aggressive drives, defences and feelings (Freud, 1923). In particular, the negotiation of the certainty of the dream description may reflect the narrator's acknowledgement of the juxtaposed and ambiguously operating secondary revision forces and his conscious effort to retrieve these inhibited memory traces. In this sense, it appears that the mitigation of dream content accuracy might reflect a conversational politeness strategy that pre-empts the unintentional communication of a detailed yet inaccurate dream description that would represent an infringement of Gricean's conversational maxims of quality, quantity, relevance and manner. Although the complete confabulation of a dream narrative might be a rare occurrence, partially confabulated dream information could easily arise within the immediate constraints of common real-life conversational situations in which a speaker feels the need to provide a coherent and socially engaging dream narrative, e.g., [20]:

"It was probably the other day. I was dreaming I was in a shop and I woke up from it and went back to sleep into it again and it was quite weird because I haven't ever worked in a shop before. People kept on coming in and asking me questions rather than buying anything and that's about all I remember really about it. People would come in one by one not ever a crowd it was always one by one and it was a fairly big shop. Yeah it was just questions about directions and stuff like that. That's about all I remember from that particular dream. I remember

waking up from certain dreams and then wanting to go back into them. Sometimes I do and sometimes I don't".

Interpretation of the third cluster

The third cluster includes the linguistic variables seeing words, first-person singular pronouns, negative emotion words and third-person singular pronouns. Seeing words generally refer to sensory-perceptual processes and sensations that exist only in the form of their inherent visual properties, such as, for example, the lexical item "light". First-person singular pronouns express the narrator's self as an individual entity and emphasise the "the self as a phenomenological self-representation schema" (Rogers, et al., 1977). Negative emotion words generally refer to negative affective concepts, for which emotions are often perceived to be evaluative judgments (Lutz & White, 1986). Third-person singular pronouns describe living entities, such as people, and typically act as anaphoric referent to a previously introduced third party (Wales, 1996); they may serve the stylistic function of avoiding noun repetition and over-topicalisation (Ehlers & Clark, 2000). The spontaneous retrieval of an authentic personal experience has often been related to negative emotional events that are thought to be more vivid than positive memories (Kensinger, Garoff-Eaton, & Schacter, 2006; Ehlers & Clark, 2000) and may revolve around the "tellability through breaching, violation, or deviation of a canonical script" (Bruner, 1991, p.12). In this sense, telling a negative emotional dream might implicitly convey its significance and relevance to the dreamer. The meaning-making and interpretation of negative events is, however, assumed to involve a greater cognitive effort and difficulty than positive events (McLean & Fournier, 2008). Overall, this cluster can be interpreted to represent *self- and other references* by emphasising the narrator's emotionally negative affect in relation to interactions with other protagonists and the visual perceptual experiences of the dream event. The spontaneous recall of negative emotional dreams also resonates with the proposition of a biological function of dreaming in which the rehearsal of the threatening dream content is assumed to lead to the evolutionary advantage of threat avoidance as a means of increasing the probability of reproductive success (Revonsuo, 2000), e.g., [11]:

"So, I went to the dentist and it was a hospital and in the waiting room the

nurse gave me a tattoo, a permanent tattoo of my appointment time on my shoulder and a message to the dentist that told him which teeth he had to take out which was supposed to be my back wisdom teeth. And she doesn't speak to the dentist and the dentist is deaf. So I go into the dentist's room and he looks at my tattoo and he reads it wrongly and he starts pulling out my front two teeth. And there's blood and it hurts and I'm screaming and he's pulling out my teeth. And then I say 'you're supposed to be taking out my back teeth not my front teeth'. And he says 'because it's the same thing, because all your teeth will move and they'll meet in the middle and you will look the same', and I was crying. And then I woke up".

Interpretation of the fourth cluster

The fourth cluster is based on the linguistic variables past tense words, insight words and causation words. The past tense typically represents specific event actions that refer to the narrator's life (Comrie, 1985). The use of the past tense also introduces a temporal and personal distance by recalling the event actions of the dream. Insight words (e.g., "know", "remember", and "believe") and causality words articulate the fundamental elements of human thinking and reasoning (Stukker et al., 2008; Verhoeven & Van Hell, 2008) and "explain the repetitive cycle of unobservable internal mental states, actions of humans and other animate agents, and observable physical states" (Kemper & Edwards, 1986, p.13). Psychological causality is most frequently introduced in discourse processes through the use of causal connectives, such as "because" and "so" (McCabe & Peterson, 1985). The use of insight words and causation words also resonates with the notion of the 'rationalisation' processes of introspective narratives (Pennebaker & King, 1999). The focus on *reasoning processes* in this cluster might reveal the narrator's autobiographical interpretation and evaluation of the dream experience by establishing a causal coherence between the self and past events (McLean & Fournier, 2008). It indicates an intuitive understanding that the content of a dream might be triggered by one's personal behavioural, cognitive and emotional experiences, rather than representing chance occurrences (Cariola, 2008). The reasoning of the thematic and emotional content also allows the rememberer to discuss the associated real-life event and dream content from a more distanced and objective perspective. Previously unacknowledged

insights, such as emotions and wishes, may be then clarified and integrated as integral parts of the re-actualised self schema and goal hierarchies (Conway & Pleydell-Pearce, 2000). This type of reasoning, by drawing causal associations between dream memories and real-life events, might assume a socially adaptive and cognitive and behaviourally assimilatory function (Conway, 2005). In summary, this cluster focuses on past tense processes that establish a causal association between the dream event and the dreamer's autobiographical history, e.g., [30]:

"Okay the only dream I really remember is about me and my dog. One day I think... it leads from I let my dog out in the garden once when she was younger. She was a rescue dog and I had to have a lot of control over her because she was very scared of people and she'd bite them. And one day when I wasn't looking she got out and she bit the postman and he wanted to prosecute me and have my dog put down which is understandable but it was a…it was a bad mistake. But that same night when I went to sleep I had a dream about me and my dog…[uh] taking my dog to an island very strange on a boat. And then me actually trying to kill my dog because I didn't want anyone else to do it. And it became a nightmare obviously because I was very upset to have to do it but [um] I couldn't kill her. I hit her over the head with a spade and lots of heavy things and she just wouldn't die. And it was just me and my dog on this very deserted little island. And me and I just remember very clearly my dog just sort of looking at me like to say why are you doing this? And that was pretty much the end of the dream. That's when I woke up".

Interpretation of the fifth cluster

The fifth cluster comprises the linguistic variables motion words, articles, spatial words and inclusion words, describing the notion of coherent movement in relation to a definite spatial context. Motion words represent a relativity dimension that is implicitly related to an existing spatiotemporal frame of reference and are typically represented through transitive verbs (e.g., "walk", "skip", and "blow") to convey the manner and cause of a movement event (Talmy 1985). In contrast, spatial

representations are generally perceived as concrete and physical sensory experiences of spatial scenes, for which lexicalised spatial representations convey the locations of objects and the relationships between these objects (Landau, 1994; Ungerleider & Mishkin, 1982). Spatial prepositions establish deictic references to the setting of a narrative that serve as "a structure into which the reminder of the story can be linked" (Rummelhart, 1975, p.214). Spatial references are also used to construct metaphors to express abstract scenarios (Lakoff & Johnson, 1980). In relation to dream narratives, spatial information often conveys an irreal spatial frame in which the dream event, protagonists and antagonists are situated (Cariola, 2008). Spatial information in dream narratives has also been perceived as a staging strategy to draw attention to the dreamer as the main protagonist of the dream event (McCarthy, 1991). Inclusion words are linked connections that increase the topical continuity between two or more entities within a spatial environment, such as the conjunction "and" (Evans & Green, 2006; Bestgen, 1998). In relation to the function of this cluster, it appears that these linguistic variables focus on the *spatial and motion processes* of the narrated dream experience. The notion of an interrelated spatial and motion dynamic was proposed by Talmy's (2000) basic motion event structure based on the assumption of "one object (Figure) that moves (Motion) along (Path) in relation to a spatial reference (Ground)" (Talmy, 2000, p. 25), in which the inclusion words narrate the coherent motion of the object through its defined spatial context. The association between motion words and spatial words in the latent structure of oral dream narratives might also reflect that the fictive motion events are being encoded spatially (Richardson & Matlock, 2005), e.g., [34]:

"Yes. It's a recurring dream happens every couple of weeks. You got…tell you what basically it's teeth falling out, it's um like eating might be a couple of different situations might be eating and then once I have taken my teeth away a couple of the teeth are still in there. In other ones I might be talking. When I go to close my mouth I shut my teeth together. They come out when they're pushed together and then if I try and like put my tongue against them to see what's happening they get pushed out. Yeah I don't know what it is, it just happens just every couple of weeks or so, just basically just teeth falling out all the time. That's more or less it".

4. Discussion

The results of the hierarchical cluster analysis identified a total of five linguistic clusters that characterise the latent structure of orally elicited dreams: 1) spoken discourse style, 2) level of detail, 3) self- and other references, 4) reasoning processes, and 5) spatial and motion processes. These linguistic clusters represent a generalised latent structure that spans the thematic and temporal boundary units of the macronarrative structures. Most interestingly, the analysis determined that these linguistic clusters reflect a dynamic interaction that functions to support the composition of a temporally, thematically and causally coherent oral dream narrative that forms part of an individual's biographical knowledge (Habermas & Bluck, 2000). In particular, it appears that the first and second dimensions focus on the accuracy mitigation of the conveyed dream content through the use of conversational discourse markers and modifiers, while the third and fourth clusters emphasise autobiographical reasoning processes that primarily attempt to make causal inferences between the rememberer's dream content and real-life experiences. The fifth cluster indicates a separate function that provides the spatial context in which the dream's events are situated. Considering the possible evolutionary advantages that have been associated with these reasoning processes (Conway & Pleydell-Pearce, 2000; Conway, 2005) and with the biological function of dreaming in particular (Revonsuo, 2000), the exploration and interpretation of manifest dream content has been perceived in the Freudian tradition as a self-reparatory function in which the rememberer might gain an improved self-understanding by confronting and overcoming emotional conflicts that come to the surface during dream interpretation.

It must, however, be taken into consideration that the results of this computer-assisted content analysis study represent merely a generalised statistical exploration of a latent linguistic structure that might not be representative of all possible linguistic variations in oral dream narratives. A few methodological problems need to be addressed to obtain an informed understanding of the results. The analysis of this study was based on a hierarchical cluster analysis, which has the advantage of enabling quantitative exploration of the linguistic structure of a given discourse type. However, it also presents the disadvantage of forcing variables to be classified within discrete clusters without necessarily acknowledging the weaker associations that might exist between

the linguistic variables (Hattie, 2002). In this sense, the latent linguistic structure proposed by cluster analysis is merely a statistical generalisation to a wider population that does not take rarer instances of linguistic patterns into account.

The cluster analysis was also based on the frequency of linguistic variables as computed by the LIWC. The LIWC generally also lacks the sensitivity to detect the contextual meaning of lexical items; therefore, some words may be wrongly categorised (Newman et al., 2008). For example, the LIWC is unable to detect the ambiguity of lexical items and the metaphorical use of polysemy, in which the meaning of semantic items often depends on their relationships with adjacent lexical items concepts and contexts (DiNardo et al., 2005; Landauer et al., 1998; Cruse, 2000). The inability to distinguish the contextual meaning of semantic items also relates to the ambiguity of grammatical forms due to their restriction to a set of words. For example, past tense words refer only to grammatical forms that unambiguously represent references to the past, e.g., "did". Complex tense constructions that combine an auxiliary with a main verb, e.g., "did go" or "were going", are thus computed as separate tenses. In the results of this study, the analysis of temporal processes has to be understood literally; present tense processes relate only to verb forms based on lexical items that indicate present tense verb forms. These present tense forms may, however, occur in combination with past tense auxiliaries.[2]

Furthermore, the LIWC taps into linguistic clusters that are based on lexical items that may co-occur in several lexical categories simultaneously. Although the linguistic categories are identical, the LIWC is biased towards covariation, and the statistical results might be skewed, to a certain extent. Conversely, Pennebaker and colleagues issued a poster that precisely outlines the lexical content of each linguistic category, eliminating the so-called "black box" that is typically associated with commercial dictionaries (Neundorf, 2002).

[2] A separate Pearson's correlation momentum coefficient, however, indicated a non-significant association between auxiliary verbs and present tense verbs, $r = .23$, $p = .15$, and past tense verbs, $r = .25$, $p = .13$, and a significant negative association between present tense verbs and past tense verbs, $r = -.74$, $p < .000$, demonstrating that the dream narratives in this study used very few complex past tense verb constructions alone.

Future research could explore how the use of a scoring system of the linguistic variables associated to each cluster might perhaps relate to individual differences of the conveyed dream content. The use of cluster analysis and the discursive function of linguistic groupings could be also applied to a large corpus of narratives of everyday memories and dreams, as an attempt to explore and interpret the cognitive processes that are particular to the narrative recall of each memory type (Cariola, in preparation).

References

Aldenderfer, M. S. & Blashfield, R. K. (1984). *Cluster analysis*. Newbury Park, CA: Sage Publications.

Bestgen, Y. (1998). Segmentation markers as trace and signal of discourse structure. *Journal of Pragmatics*, *29*, 753-763.

Biber, D. (1988). *Variations across speech and writing*. Cambridge, UK: Cambridge University Press.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge, UK: Cambridge University Press.

Blashfield, R. K. (1976). Mixture model tests of cluster analysis: Accuracy of four agglomerative hierarchical methods. *Psychological Bulletin*, *83*, 377-388.

Brown, G., & Yule, G. (1983). *Discourse Analysis*. Cambridge, UK: Cambridge University Press.

Bruner, J. (1991). The Narrative Construction of Reality. *Critical Inquiry, 18*, 1-21.

Cariola, L. A. (in preparation). Linguistic differences in narratives of everyday memories and dreams.

Cariola, L. A. (2008). A structural and functional analysis of dream narratives. *Dreaming, 18*, 16-26.

Cariola, L. A. (2011). Analysing the latent linguistic structure of American-English and German dream narratives. *The 28th Annual Conference of the International Association for the Study of Dreams - Dreams and Cultural Diversity*, Kerkrade, Netherlands.

Cariola, L. A., O'Connor, D. B., & Conway, M. A. (2009). Linguistic structures in true, plausible false and false autobiographical memories. *Postgraduate Conference, Institute of Psychological Sciences*, University of Leeds.

Comrie, B. (1976). *Aspect*. Cambridge: Cambridge University Press.

Comrie, B. (1985). *Tense*. Cambridge: Cambridge University Press.

Conway, M. A. (2005). Memory and the self. *Journal of Memory and Language, 53,* 594-628.

Conway, M. A., & Bekerian, D. A. (1987). Organisation in autobiographical memory. *Memory and Cognition, 15*, 119-132.

Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological Review, 107*, 261-288.

Cruse, A. (2000). *Meaning in Language: an introduction to semantics and pragmatics*. Oxford University Press, Oxford, 2000.

DiNardo, A. C, Michael F., Schober, M. F., & Stuart, J. (2005). Chair and couch discourse: A study of visual copresence in psychoanalysis. *Discourse Processes, 40*, 209-238.

Edelbrock, C. (1979). Mixture model tests of hierarchical clustering algorithms: the problem of classifying everybody. *Multivariate Behavioral Research, 14*, 367-384.

Ehlers, A., & Clark, D. M. (2000). A cognitive model of posttraumatic stress disorder. *Behaviour Research and Therapy, 38*, 319-345.

Evans, V., & Green, M. (2006). *Cognitive linguistics: An introduction*. Edinburgh, UK: Edinburgh University Press.

Everitt, B. S. (1979). Unresolved problems in cluster analysis. *Biometrics, 35*, 169-181.

Fraser, B., & Malamud-Makowski, M. (1996). English and Spanish contrastive discourse markers. *Language Science, 18*, 863-881.

Freud, S. (2001). *The Interpretation of Dreams*. *SE 4*. London, UK: Hogarth Press. (Original work published in 1900)

Freud, S. (2001). *The Ego and the Id*, *SE 19*. London, UK: Hogarth Press. (Original work published in 1923)

Goldsmith, M., Koriat, A., & Weinberg-Eliezer, A. (2002). Strategic regulation of grain size memory reporting. *Journal of Experimental Psychology: General*, 131, 73–95.

Goldsmith M., Koriat A., & Pansky, A. (2005). Strategic regulation of grain size in memory reporting over time. *Journal of Memory and Language, 52*, 505-525.

Habermas, T. & Bluck, S. (2000). Getting a life: The emergence of the life story in adolescence. *Psychological Bulletin*, *126*, 748-769.

Halliday, M. A. K. (1994). Spoken and written modes of meaning. In D. Graddol & O. Boyed-Barrett (Eds.), *Media texts, authors and readers: a reader* (pp.51-73). Clevedon, UK: Multilingual Matters.

Hattie, J. A. (2002). Schools Like Mine: Cluster Analysis of New Zealand Schools. *Technical Report 14, Project asTTle*. University of Auckland.

Holmes, J. (1995). *Women, men and politeness*. London, UK: Longman.

Holmes, J. (1990). Hedges and boosters in women's and men's speech. *Language and Communication*, *10*, 185-205.

Holland, S. M. (2006). *Cluster analysis.* www.uga.edu/strata/software/pdf/clusterTutorial.pdf Retrieved on the 30th November 2010

Izutsu, M. N. (2008). Contrast, concessive, and corrective: Toward a comprehensive study of opposition relations. *Journal of Pragmatics, 40*, 646–675.

Kahn, J. H., Tobin, R. M., Massey, A. E., & Anderson, J. A. (2007). Measuring emotional expression with the Linguistic Inquiry and Word Count. *American Journal of Psychology, 120*, 263-286.

Kemper, S., & Edwards L. (1986). Children's expression of causality and their construction of narratives. *Topics in Language Disorders, 7*, 11-20.

Kensinger, E. A., Garoff-Eaton, R. J., & Schacter, D. L. (2006). Memory for specific visual details can be enhanced by negative arousing content. *Journal of Memory and Language*, *54*, 99-112.

Labov, W., & Waletzky, J, (1967). Narrative analysis: Oral versions of personal experience. In J. Helms (Ed.) *Essays on the Verbal and Visual Arts: Proceedings of the 1966 Annual Spring Meeting of the American Ethnological Society*. Seattle: University of Washington Press.

Lakoff, J. & Johnson, M. (1980). *Metaphors We Live By*. Chicago: University of Chicago Press.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes, 25*, 259-284.

Lutz, C., & White, G. M. (1986). The anthropology of emotions. *Annual Review of Anthropology, 15*, 405-436.

McCabe, A., & Peterson, C. (1985). A naturalistic study of the production of causal connectives by children. *Journal of Child Language*, *12*, 145-159.

McCarthy, M. (1991). *Discourse analysis for language teachers*. Cambridge, United

Kingdom: Cambridge University Press.

McNally, L., & Kennedy, C. (2008). *Adjectives and adverbs: Syntax, semantics, and discourse (Oxford Studies in Theoretical Linguistics 19)*. Oxford, UK: Oxford University Press.

McLean, K. C., & Fournier, M. A. (2008). The content and processes of autobiographical reasoning in narrative identity. *Journal of Research in Personality, 42*, 527-545.

Mulac, A., & Lundell, T. L. (1994). Effects of gender-linked language differences in adults' written discourse: Multivariate tests of language effects. *Language & Communication, 14*, 299–309.

Neundorf, K. A. (2002). The Content Analysis Guidebook. London, UK: Sage Publication.

Newman, M. L., Groom, C. J., Handelman, L. D., & Pennebaker, J. W. (2008). Gender differences in language use: An analysis of 14.000 text samples. *Discourse Processes, 45*, 211-236.

Ochs, E., & Capps, L. (1997). Narrative authenticity. *Journal of Narrative and Life History, 7*, 83-89.

Pennebaker, J. W., & King, L. A. (1999). Linguistic styles: Language use as an individual difference. *Journal of Personality & Social Psychology, 6*,1296–1312.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). *Linguistic Inquiry and Word Count (LIWC): LIWC2001*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Revonsuo, A. (2000). The interpretation of dreams: An evolutionary hypothesis of the function of dreaming. *Behavioral and Brain Sciences*, *23*, 793-1121.

Richardson, D., & Matlock, T. (2005). The integration of figurative language and static depictions: An eye movement study of fictive motion. *Cognition, 102*, 129-138.

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology, 35*, 677-88.

Rummelhart, D. E. (1975). Notes on schema for stories. InD. G. Bobdrow, & A. Collins (Eds.), *Representation and understanding, studies in cognitive science* (pp.). New York, NY: Academic Press.

Schourup, L. (1999). Discourse markers. *Lingua*, *107*, 227-265.

Singer, J. A., & Bluck, S. (2001). New perspectives on autobiographical memory: The integration of narrative processing and autobiographical reasoning. *Review of General Psychology*, 5, 91-99.

Snyder, F. (1970). The phenomenology of dreaming. In L. Madow & L. Snow (Eds.), *The psychodynamic implications of the physiological studies on dreams* (pp. 124-151). Springfield, IL: Thomas.

Stukker, N., Sanders T., & Verhagen A. (2008). Causality in verbs and in discourse connectives. Converging evidence of cross-level parallels in Dutch linguistic categorization. *Journal of Pragmatics, 40*, 1296-1322.

Tagliamonte, S. (2005). So who? Like how? Just what? Discourse markers in the conversations of Young Canadians. *Journal of Pragmatics*, *37*, 1896-1915.

Talmy, L. (1985). Lexicalization patterns: Semantic structure in lexical forms. In T. Shoepen (Ed.), *Language typology and lexical description: Vol. 3. Grammatical categories and the lexicon* (pp. 36-149). Cambridge: Cambridge University Press.

Talmy, L. (2000). *Toward a cognitive semantics: Vol. II: Typology and process in concept structuring*. Cambridge, MA: MIT Press.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology, 29*, 24-54.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. Ingle, M. Goodale, & R. Mansfield (Eds.). Analysis of Visual Behaviour (pp.263-299). Cambridge, MA: MIT Press.

Verhoeven, L., & Van Hell, J. G. (2008). From knowledge representation to writing text: A developmental perspective. *Discourse Processes, 45*, 387-405.

Wales, K. (1996). *Personal Pronouns in Present-day English*. Cambridge: University Press.

Williams, W. T., Clifford, H. T., & Lance, G. N. (1971a). Group size dependence: a rationale for choice between numerical classification. *The Computer Journal 14*, 157-162.

Wilson, A. (2009). Speech, writing and discourse type. In J. Culpeper, T. McEnery, F. Katamba, & R. Wodak (Eds.), *English Language: Descriptions, Variation and Context* (pp.425-438). Basingstoke, UK: Palgrave MacMillan.

Appendix

Table 1 – Descriptive statistics of linguistic variables in orally elicited dream narratives.

	Maximum	Minimum	Mean	SD
Linguistic processes				
Word count	46.00	646.00	216.26	116.31
1st person singular pronouns	8.84	8.89	8.84	3.23
3rd person singular pronouns	1.26	.34	1.26	1.86
Past tense	8.88	9.74	8.88	3.81
Present tense	5.78	5.38	5.78	3.29
Adverbs	7.06	6.81	7.06	2.70
Quantifiers	2.29	1.96	2.29	1.62
Positive emotions	1.90	1.77	1.90	1.20
Negative emotions	1.32	.86	1.32	1.28
Insight	2.44	2.39	2.44	1.36
Causation	1.06	1.00	1.06	.90
Tentativeness	2.61	2.07	2.61	1.97
Certainty	1.11	.92	1.11	1.07
Inclusion	9.17	8.30	9.17	3.83
Exclusion	3.24	3.14	3.24	1.78
Sight	1.03	.56	1.03	1.38
Motion	2.56	2.67	2.56	1.62
Space	7.67	7.84	7.67	2.85

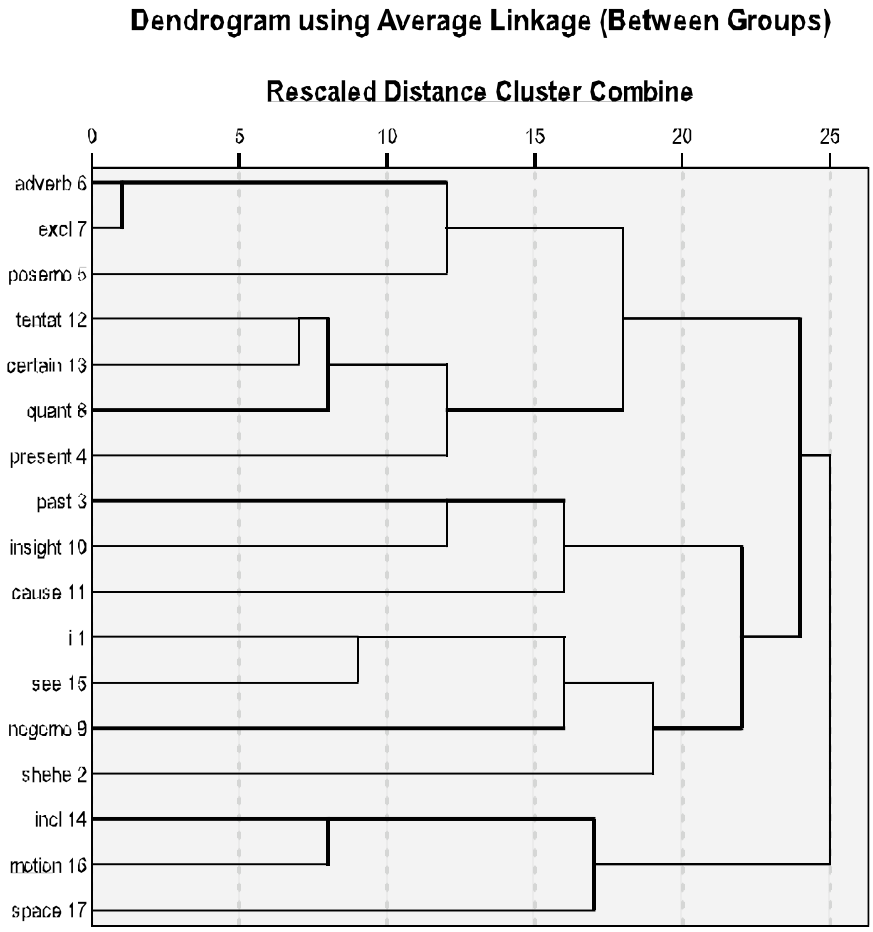

Figure 1 – Cluster dendogram using an average linkage with a Pearson correlation product momentum coefficient of linguistic variables in orally elicited dream narratives.